文章编号:1000-5862(2014)02-0176-04

两参数 Lomax 分布中参数的区间估计和假设检验

龙 兵
(荆楚理工学院数理学院,湖北 荆门 448000)

摘要: 研究了两参数 Lomax 分布参数的区间估计和假设检验问题. 分别在形状参数和尺度参数已知的情形下,给出了尺度参数、形状参数的置信区间和假设检验的拒绝域以及 p 分位数的置信区间,并运用随机模拟的方法分别对 2 个参数进行了统计分析.

关键词: Lomax 分布; χ^2 分布; 区间估计; 假设检验
中图分类号: O 212.2 **文献标志码:** A

0 引言

Lomax 分布是数理统计中一种重要分布. 由于该分布的失效率是单调递减和单调递增的,并且在生命科学和可靠性工程学等方面的寿命试验数据处理中起着重要作用,因此有许多统计学者对此分布进行了研究,得到了一些结果. 文献[1-8]研究了不同损失函数下,当尺度参数已知时形状参数的 Bayes 估计,文献[9]研究了两参数 Lomax 分布次序统计量的性质和渐近分布.

已有的研究成果对于参数的 Bayes 估计问题都是假设尺度参数已知,对形状参数进行点估计. 而对尺度参数进行区间估计和假设检验的研究文献较少见. 本文将分别对形状参数和尺度参数进行区间估计和假设检验,因此该研究就显得尤为重要.

两参数 Lomax 分布的概率密度函数为

$$f(x)=\frac{\theta}{\lambda}\left(1+\frac{x}{\lambda}\right)^{-(\theta+1)},x>0,\lambda>0,\theta>0,\quad(1)$$

其分布函数为

$$F(x)=1-(1+x/\lambda)^{-\theta},x>0,\lambda>0,\theta>0,$$

其中 λ 为尺度参数,θ 为形状参数.

由(1)式得,当 $\theta>1$ 时,Lomax 分布的期望 $E(X)=\lambda/(\theta-1)$,当 $\theta>2$ 时,其方差为

$$D(X)=\lambda^2\theta/[(\theta-1)^2(\theta-2)],$$

变异系数为

$$C_v(X)=\sqrt{D(X)}/E(X)=\sqrt{\theta/(\theta-2)}.$$

在全样本下对参数 λ,θ 的经典点估计可以通过矩估计法和极大似然法得到,这里不作研究.

1 Lomax 分布中参数 λ,θ 及 p 分位数的区间估计

假设 $Y\sim U(0,1)$,从总体 $U(0,1)$ 中抽取容量为 n 的样本 $(Y_1,Y_2,\cdots,Y_n)$,构造统计量 $p(n)=Y_1\cdot Y_2\cdots Y_n$,则有如下引理.

引理 1[10] $p(n)$ 的分布函数为

$$F_{p(n)}(z)=\begin{cases}0,z<0,\\ z\left[1-\ln z+\dfrac{(-\ln z)^2}{2!}+\cdots+\dfrac{(-\ln z)^{n-1}}{(n-1)!}\right],0\leqslant z<1,\\ 1,z\geqslant 1,\end{cases}\quad(2)$$

$p(n)$ 的密度函数为

$$f_{p(n)}(z)=\frac{(-1)^{n-1}}{(n-1)!}(\ln z)^{n-1},0<z<1.\quad(3)$$

证 只需要证明 $p(n)$ 的密度函数为(3)式即可. 对(3)式积分就可得(2)式.

下面用归纳法进行证明.

当 $k=1$ 时,$p(1)=Y_1$,其密度函数为 $f_{p(1)}(z)=1,0<z<1$,结论显然成立.

假设当 $k=n$ 时,结论成立,即

$$f_{p(n)}(z)=\frac{(-1)^{n-1}}{(n-1)!}(\ln z)^{n-1},0<z<1.$$

接下来证明当 $k=n+1$ 时,结论也是成立的.

因为 $p(n+1)=U_1U_2\cdots U_nU_{n+1}=p(n)U_{n+1}$,且

收稿日期:2013-12-20
基金项目:湖北省教育厅重点科研课题(D20134301)资助项目.
作者简介:龙 兵(1973-),男,湖北荆门人,副教授,主要从事数理统计方面的研究.

$U_{n+1}\sim U(0,1)$，$p(n)$ 与 U_{n+1} 独立，所以 $p(n+1)$ 的分布函数为

$$F_{p(n+1)}(z)=P(p(n)U_{n+1}\leqslant z)=\iint_{xy\leqslant z,0<x,y<1}\frac{(-\ln y)^{n-1}}{(n-1)!}\mathrm{d}x\mathrm{d}y.$$

当 $z<0$ 时，$F_{p(n+1)}(z)=0$；当 $z\geqslant 1$ 时，$F_{p(n+1)}(z)=1$；当 $0\leqslant z<1$ 时，

$$F_{p(n+1)}(z)=1-\int_z^1\frac{(-\ln y)^{n-1}}{(n-1)!}\mathrm{d}y\int_{z/y}^1\mathrm{d}x=$$

$$1-\int_z^1\frac{(-\ln y)^{n-1}}{(n-1)!}\left(1-\frac{z}{y}\right)\mathrm{d}y=$$

$$1-\left[\int_z^1\frac{(-\ln y)^{n-1}}{(n-1)!}\mathrm{d}y-\frac{(-1)^n}{n!}z(\ln z)^n\right].$$

对上式两端关于 z 求导，得

$$f_{p(n+1)}(z)=\frac{\mathrm{d}F_{p(n+1)}(z)}{\mathrm{d}z}=-\left[-\frac{(-\ln z)^{n-1}}{(n-1)!}-\frac{(-1)^n}{n!}\cdot\frac{\mathrm{d}}{\mathrm{d}z}z(\ln z)^n\right]=\frac{(-\ln z)^{n-1}}{(n-1)!}+$$

$$\frac{(-1)^n}{n!}\left[(\ln z)^n+n(\ln z)^{n-1}\right]=\frac{(-1)^n}{n!}(\ln z)^n.$$

定理 1　设随机变量 X 服从 $P(n)$ 分布，令 $Y=-2\ln X$，则 $Y\sim\chi^2(2n)$.

证　$F_Y(y)=P(Y\leqslant y)=P(-2\ln X\leqslant y)=P(2\ln X\geqslant -y)=P(X\geqslant \mathrm{e}^{-y/2})=\int_{\mathrm{e}^{-y/2}}^1\frac{(-\ln x)^{n-1}}{(n-1)!}\mathrm{d}x$，

$$f_Y(y)=F'_Y(y)=\frac{y^{n-1}}{2^n(n-1)!}\mathrm{e}^{-y/2},0<y<\infty.$$

这正是 $\chi^2(2n)$ 分布的概率密度函数.

设 $X_1,X_2,\cdots,X_n$ 是来自两参数 Lomax 分布(1)的独立同分布随机样本.

令 $U_i=1-(1+X_i/\lambda)^{-\theta}$，$V_i=1-U_i$，$i=1,2,\cdots,n$，则 V_i 服从均匀分布 $U(0,1)$. 构造随机变量

$$p=V_1V_2\cdots V_n=\prod_{i=1}^n\left(1+\frac{X_i}{\lambda}\right)^{-\theta},$$

由引理 1 知 $p\sim p(n)$，则

$$Y=-2\ln p=-2\sum_{i=1}^n\ln\left(1+\frac{X_i}{\lambda}\right)^{-\theta}=2\theta\sum_{i=1}^n\ln\left(1+\frac{X_i}{\lambda}\right),$$

由定理 1 知 $Y\sim\chi^2(2n)$.

下面讨论参数及分位数的区间估计问题.

1.1　当尺度参数 λ 已知时，形状参数 θ 的区间估计

取枢轴量 $Y=2\theta\sum_{i=1}^n\ln(1+X_i/\lambda)$，$\forall\alpha\in(0,1)$，对于给定的置信水平 $1-\alpha$，考虑

$$P\left(\chi^2_{\alpha/2}(2n)\leqslant 2\theta\sum_{i=1}^n\ln\left(1+\frac{X_i}{\lambda}\right)\leqslant\chi^2_{1-\alpha/2}(2n)\right)=1-\alpha,$$

即

$$P\left(\frac{\chi^2_{\alpha/2}(2n)}{2\sum_{i=1}^n\ln\left(1+\frac{X_i}{\lambda}\right)}\leqslant\theta\leqslant\frac{\chi^2_{1-\alpha/2}(2n)}{2\sum_{i=1}^n\ln\left(1+\frac{X_i}{\lambda}\right)}\right)=1-\alpha.$$

因此，θ 的置信水平 $1-\alpha$ 的置信区间为

$$\left[\frac{\chi^2_{\alpha/2}(2n)}{2\sum_{i=1}^n\ln(1+X_i/\lambda)},\frac{\chi^2_{1-\alpha/2}(2n)}{2\sum_{i=1}^n\ln(1+X_i/\lambda)}\right].\tag{4}$$

1.2　当形状参数 θ 已知时，尺度参数 λ 的区间估计

仍取枢轴量 $Y=2\theta\sum_{i=1}^n\ln(1+X_i/\lambda)$，对于置信水平 $1-\alpha$，则

$$P\left(\chi^2_{\alpha/2}(2n)\leqslant 2\theta\sum_{i=1}^n\ln(1+X_i/\lambda)\leqslant\chi^2_{1-\alpha/2}(2n)\right)=1-\alpha.$$

由于 Y 是 λ 的严格单调递减函数，解不等式

$$\chi^2_{\alpha/2}(2n)\leqslant 2\theta\sum_{i=1}^n\ln(1+X_i/\lambda)\leqslant\chi^2_{1-\alpha/2}(2n)$$

可以得到

$$\hat{\lambda}_L(X_1,X_2,\cdots,X_n)\leqslant\lambda\leqslant\hat{\lambda}_U(X_1,X_2,\cdots,X_n).$$

因此，λ 的置信水平为 $1-\alpha$ 的置信区间为

$$[\hat{\lambda}_L(X_1,X_2,\cdots,X_n),\hat{\lambda}_U(X_1,X_2,\cdots,X_n)].$$

事实上，对任给的正实数 k，方程

$$h(\lambda)=2\theta\sum_{i=1}^n\ln(1+X_i/\lambda)-k=0$$

都有唯一解. 此因 $\lambda\to+\infty$，$2\theta\sum_{i=1}^n\ln(1+X_i/\lambda)-k\to -k$，$\lambda\to 0^+$，$2\theta\sum_{i=1}^n\ln(1+X_i/\lambda)-k\to+\infty$，且 $2\theta\sum_{i=1}^n\ln(1+X_i/\lambda)-k$ 是 λ 的严格单调递减函数.

1.3　分位数的区间估计

由于分布函数 $F(x)=1-(1+x/\lambda)^{-\theta}$，$\forall p\in(0,1)$，当 $F(x_p)=p$ 时，解得 Lomax 分布的 p 分位数为 $x_p=\lambda[(1-p)^{-1/\theta}-1]$，可以看到 x_p 是 λ 的单调递增函数，因此，当 θ 已知时，p 分位数 x_p 的 $1-\alpha$ 置信区间为

$$[\lambda_L((1-p)^{-1/\theta}-1),\lambda_U((1-p)^{-1/\theta}-1)].$$

又因为 x_p 是 θ 的单调递减函数，因此，当 λ 已知时，p 分位数 x_p 的 $1-\alpha$ 置信区间为

$$[\lambda((1-p)^{-1/\theta_U}-1),\lambda((1-p)^{-1/\theta_L}-1)],\tag{5}$$

其中

$$\hat{\theta}_L=\frac{\chi^2_{\alpha/2}(2n)}{2\sum_{i=1}^n\ln(1+X_i/\lambda)},\hat{\theta}_U=\frac{\chi^2_{1-\alpha/2}(2n)}{2\sum_{i=1}^n\ln(1+X_i/\lambda)}.$$

2 关于参数 λ,θ 的假设检验

2.1 当 λ 已知时,θ 的假设检验

首先讨论 $H_0:\theta \leqslant \theta_0, H_1:\theta > \theta_0$. 取 $Y = 2\theta \cdot \sum_{i=1}^{n} \ln(1 + X_i/\lambda)$ 作为检验统计量,则 Y 是 θ 的严格单调递增函数.

当原假设 H_0 成立时,

$$\left\{2\theta_0 \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda}\right) \leqslant k\right\} \subset \left\{2\theta \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda}\right) \leqslant k\right\},$$

取

$$P\left\{2\theta_0 \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda}\right) \leqslant \chi_\alpha^2(2n)\right\} \leqslant P\left\{2\theta \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda}\right) \leqslant \chi_\alpha^2(2n)\right\} = \alpha,$$

所以检验的拒绝域

$$W = \left\{(X_1, X_2, \cdots, X_n) \,\middle|\, 2\theta_0 \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda}\right) \leqslant \chi_\alpha^2(2n)\right\}.$$

用类似的方法对 $H_0:\theta \geqslant \theta_0, H_1:\theta < \theta_0$,检验的拒绝域取为

$$W = \left\{(X_1, X_2, \cdots, X_n) \,\middle|\, 2\theta_0 \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda}\right) \geqslant \chi_{1-\alpha}^2(2n)\right\}.$$

对 $H_0:\theta = \theta_0, H_1:\theta \neq \theta_0$ 检验的拒绝域取为

$$W = \left\{(X_1, X_2, \cdots, X_n) \,\middle|\, 2\theta_0 \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda}\right) \leqslant \chi_{\alpha/2}^2(2n) \text{ 或 } 2\theta_0 \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda}\right) \geqslant \chi_{1-\alpha/2}^2(2n)\right\}.$$

2.2 当 θ 已知时,λ 的假设检验

对于假设检验 $H_0:\lambda \leqslant \lambda_0, H_1:\lambda > \lambda_0$. 取 $Y = 2\theta \sum_{i=1}^{n} \ln(1 + X_i/\lambda)$ 作为检验统计量,显然 Y 是 λ 的严格单调递减函数.

当 H_0 成立时,有

$$\left\{2\theta \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda_0}\right) \geqslant k\right\} \subset \left\{2\theta \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda}\right) \geqslant k\right\},$$

取

$$P\left\{2\theta \sum_{i=1}^{n} \ln(1 + X_i/\lambda_0) \geqslant \chi_{1-\alpha}^2(2n)\right\} \leqslant P\left\{2\theta \sum_{i=1}^{n} \ln(1 + X_i/\lambda) \geqslant \chi_{1-\alpha}^2(2n)\right\} = \alpha,$$

所以检验的拒绝域为

$$W = \left\{(X_1, X_2, \cdots, X_n) \,\middle|\, 2\theta \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda_0}\right) \geqslant \chi_{1-\alpha}^2(2n)\right\}.$$

类似地,对假设检验问题 $H_0:\lambda \geqslant \lambda_0, H_1:\lambda < \lambda_0$. 检验的拒绝域取为

$$W = \left\{(X_1, X_2, \cdots, X_n) \,\middle|\, 2\theta \sum_{i=1}^{n} \ln\left(1 + \frac{X_i}{\lambda_0}\right) \leqslant \chi_\alpha^2(2n)\right\}.$$

对检验问题 $H_0:\lambda = \lambda_0, H_1:\lambda \neq \lambda_0$. 检验的拒绝域取为

$$W = \left\{(X_1, X_2, \cdots, X_n) \,\middle|\, 2\theta \sum_{i=1}^{n} \ln(1 + X_i/\lambda_0) \leqslant \chi_{\alpha/2}^2(2n) \text{ 或 } 2\theta \sum_{i=1}^{n} \ln(1 + X_i/\lambda_0) \geqslant \chi_{1-\alpha/2}^2(2n)\right\}.$$

对于 p 分位数的假设检验问题,也可以用类似的方法得到检验的拒绝域,这里不再论述.

3 随机模拟及实例分析

类似于文献[11],下面用随机模拟来验证统计方法的优劣. 具体步骤如下:

Step 1 产生1组容量为 n 的服从 $U(0,1)$ 的相互独立随机样本 $Y_1, Y_2, \cdots, Y_n$;

Step 2 给定 λ,θ 的值,令

$$x_i = \lambda\left[(1 - Y_i)^{-1/\theta} - 1\right], i = 1, 2, \cdots, n,$$

则 $x_1, x_2, \cdots, x_n$ 是服从两参数 Lomax 分布的样本;

Step 3 给定 $p = 0.4$ 及取 $\alpha = 0.05$,由(4)式和(5)式可以得到 θ 和 x_p 置信区间,从而能够判断所得区间是否包含 θ 和 x_p 的真值.

以上步骤随机模拟 1 000 次,计算不同的样本容量 n 和参数 λ,θ 下,参数 θ 及0.4分位数的覆盖率及上、下限均值. 模拟结果列于表1中.

表1 覆盖率及上、下限均值($\alpha = 0.05$)

λ	θ 的真值	n	θ 的覆盖率	$x_{0.4}$ 的覆盖率	下限均值	上限均值
1.2	3.0	10	0.950	0.946	1.595 0	5.682 6
		12	0.957	0.951	1.705 4	5.413 3
1.2	2.8	10	0.945	0.949	1.493 3	5.320 2
		12	0.947	0.944	1.605 9	5.097 6
1.2	2.6	10	0.963	0.962	1.408 6	5.018 6
		12	0.946	0.956	1.424 8	4.522 6
1.0	3.0	10	0.944	0.945	1.588 9	5.660 8
		12	0.959	0.959	1.698 6	5.391 6
1.0	2.8	10	0.946	0.957	1.482 3	5.281 1
		12	0.953	0.962	1.568 6	4.978 9
1.0	2.6	10	0.953	0.949	1.409 0	5.019 8
		12	0.951	0.952	1.467 8	4.659 2

由表1可知,对于给定的置信水平0.95,覆盖率跟0.95非常接近. 另外,随着样本量 n 的增大,下限均值增大和上限均值减小.

使用 Matlab 软件,给定 $\theta = 3, \lambda = 1$ 通过 Monte-Carlo 方法随机产生 1 组容量为 14 且服从 Lomax 分布(1)的样本,按照从小到大的顺序排列

如下:

0.079 8,0.084 8,0.136 7,0.160 1,0.183 7,

0.207 6,0.235 8,0.294 4,0.340 3,0.380 8,

0.485 6,0.684 9,1.566 1,1.709 6.

经过统计分析得到相关结论如表 2 和表 3.

表 2　当 $\lambda = 1$ 时,分位数的区间估计、θ 的区间估计和假设检验($\alpha = 0.05$)

θ 的置信区间	$x_{0.4}$ 置信区间	$x_{0.6}$ 置信区间	$H_0: \theta \leqslant 2$ $H_1: \theta > 2$	$H_0: \theta \geqslant 4.5$ $H_1: \theta < 4.5$	$H_0: \theta = 3.5$ $H_1: \theta \neq 3.5$
[1.617,4.696]	[0.115,0.372]	[0.215,0.762]	接受 H_0	拒绝 H_0	接受 H_0

表 3　当 $\theta = 3$ 时,分位数的区间估计、λ 的区间估计和假设检验($\alpha = 0.05$)

λ 的置信区间	$x_{0.4}$ 置信区间	$x_{0.6}$ 置信区间	$H_0: \lambda \leqslant 0.6$ $H_1: \lambda > 0.6$	$H_0: \lambda \geqslant 1.2$ $H_1: \lambda < 1.2$	$H_0: \lambda = 0.5$ $H_1: \lambda \neq 0.5$
[0.545,2.136]	[0.101,0.397]	[0.195,0.763]	拒绝 H_0	接受 H_0	拒绝 H_0

从表 2 和表 3 中的数据可以看到,参数真值落在置信区间的中部. 随着 p 值的增大,分位数 x_p 的置信区间长度也随之增大,这与 Lomax 分布的密度函数是单调递减的性质相符合.

4　参考文献

[1] 肖小英,任海平. 熵损失函数下两参数 Lomax 分布形状参数的 Bayes 估计 [J]. 数学的实践与认识,2010,40(5):227-230.

[2] 周明元. 对称熵损失函数下两参数 Lomax 分布形状参数的 Bayes 估计 [J]. 统计与决策,2010(12):161-162.

[3] 王琪,任海平. NA 样本下两参数 Lomax 分布形状参数的经验 Bayes 检验 [J]. 统计与决策,2010(17):8-10.

[4] 姚惠,谢林. 不同损失下 Lomax 分布形状参数的 Bayes 估计 [J]. 数学杂志,2011,31(6):1131-1135.

[5] 姚惠. Linex 损失下 Lomax 分布形状参数的 Bayes 估计 [J]. 统计与决策,2011(16):173-175.

[6] 姚惠. 熵损失函数下 Lomax 分布形状参数的 Bayes 估计 [J]. 遵义师范学院学报,2011,13(6):107-109.

[7] 芦凌飞. 刻度平方误差损失下 Lomax 分布形状参数的 Bayes 估计 [J]. 商丘师范学院学报,2012,28(6):38-40.

[8] 余慧敏. 复合 linex 对称损失下 Lomax 分布参数的 Bayes 估计 [J]. 广东海洋大学学报,2013,33(4):87-89.

[9] 龙兵. 两参数 Lomax 分布次序统计量的性质和渐近分布 [J]. 兰州交通大学学报,2013,32(4):164-167.

[10] 潘高田,王保恒,陈春良,等. 艾拉姆咖分布小样本区间估计和检验问题研究 [J]. 数理统计与管理,2009,28(3):468-472.

[11] 龙兵. 缺失数据样本下艾拉姆咖分布的参数估计 [J]. 江西师范大学学报: 自然科学版,2013,37(1):16-19.

The Interval Estimation and Hypothesis Test of the Parameters from Lomax Distribution

LONG Bing

(Department of Mathematics and Physics,Jingchu University of Technology,Jingmen Hubei 448000,China)

Abstract: Interval estimation and hypothesis test of the parameters from Lomax distribution with two parameters are discussed. Confidence intervals and hypothesis tests of scale and shape parameters are given when the shape parameter or scale parameter is known. In addition,p quantile is also obtained about Lomax distribution. In the end,statistical analysis of two parameters are carried out by means of Monte Carlo simulation.

Key words: Lomax distribution; Chi-quare distribution; interval estimation; hypothesis test

(责任编辑: 曾剑锋)